MENTORBOT – DATA SCIENCE EDUCATION CHATBOT FOR INTERACTIVE LEARNING AND PROBLEM SOLVING

CHODAVARAPU PRAVEEN KUMAR[1]

1.Student, Dept. of Computer Science and Engineering (CSE), Usha Rama College of Engineering and Technology. Near Gannavaram Ungutur Mandalam, Telaprolu, Andhra Pradesh

Dr. S.M ROY CHOUDRI[2]

2.Professor, Dept. of Computer Science and Engineering (CSE), Usha Rama College of Engineering and Technology. Near Gannavaram Ungutur Mandalam, Telaprolu, Andhra Pradesh.

Dr. BEJAGAM PRAVEENKUMAR[3]

3.M. Tech Coordinator Dept. of Computer Science and Engineering (CSE), Usha Rama College of Engineering and Technology. Near Gannavaram Ungutur Mandalam, Telaprolu, Andhra Pradesh.

ABSTRACT

The demand for data science professionals is projected to grow by 36% between 2021 and 2031, creating an urgent need for effective and scalable educational tools. While generative AI chatbots like ChatGPT are being explored in educational settings, their application has yielded mixed results. Studies in programming courses show that while students find these general-purpose tools helpful for clarification, they are often perceived as unreliable due to inaccuracies and hallucinations, and their use does not yield quantitatively significant improvements in skills like reflective thinking. These systems lack a curated knowledge base and a structured pedagogical framework, making them more of a reactive information source than a guided learning partner. This paper proposes **MentorBot**, a domain-specific conversational agent for data science education. We propose a **Modular Cognitive Pedagogical Agent (MCPA)** architecture, implemented using Google Dialogflow. MentorBot is designed with a curated knowledge base for data science and machine learning, ensuring accuracy and relevance. It integrates pedagogical question-answering with social dialogue to create an engaging, human-like interaction. By providing a controlled, accurate, and interactive learning environment, MentorBot aims to overcome the limitations of generalist AI tools and provide a more effective and reliable educational assistant for aspiring data scientists.

Keywords: Educational Chatbot, Data Science Education, Cognitive Computing, Google Dialogflow, Natural Language Processing, Pedagogical Agent, Interactive Learning, Question Answering System.

1.INTRODUCTION

Cognitive computing has grown significantly in recent years, increasing both research and commercial interest in intelligent systems. Conversational agents, commonly known as chatbots, have evolved from simple rule-based programs into advanced systems that use Natural Language Understanding (NLU) and machine learning techniques. These technologies enable chatbots to maintain flexible and meaningful conversations with users. Many organizations now use chatbots to answer frequently asked questions and provide automated assistance. A Natural Language Interface (NLI) allows users to interact with systems in a more intuitive and efficient way compared to traditional search interfaces. In education, cognitive computing systems can enhance student learning by providing personalized assistance and instant feedback. Chatbots encourage interactive learning, improve student engagement, and help teachers manage learning materials

effectively. Compared with traditional e-learning platforms, chatbot-based learning systems offer greater interactivity and accessibility.

Inspired by these advantages, this study focuses on developing an intelligent chatbot to support data science education and improve students' learning experiences.

2.LITERATURE REVIEW

1. Ronneberger et al. (2015) proposed deep learning architectures that demonstrated the effectiveness of neural networks in solving complex computational problems through automated learning from large datasets. Their work highlighted the potential of AI-based models to process complex inputs and generate accurate outputs, which later influenced

the development of intelligent conversational systems used in educational environments.

2. Colace et al. (2018) developed an **AI-based chatbot for e-learning environments** that allowed students to interact with a virtual assistant to obtain course-related information. The study showed that conversational interfaces improve accessibility and engagement in online learning systems. However, the chatbot primarily focused on answering predefined questions and lacked advanced problem-solving capabilities.

3. Winkler and Söllner (2018) conducted a study on the **design of educational chatbots** and examined their effectiveness in improving student interaction and learning support. Their research revealed that chatbots can provide instant responses to student queries and reduce the workload on instructors. Despite these advantages, their system lacked adaptive learning features and personalized feedback mechanisms.

4. Okonkwo and Ade-Ibijola (2021) reviewed the **application of chatbots in education** and concluded that AI-driven conversational agents significantly enhance learning experiences by offering real-time guidance and tutoring. The authors also emphasized the importance of integrating Natural Language Processing and machine learning techniques to improve chatbot understanding and response accuracy.

5. Recent research on data science learning assistants has explored the integration of chatbots with machine learning frameworks to guide students in coding tasks, data analysis, and algorithm understanding. These systems provide interactive explanations, coding suggestions, and quiz-based learning modules. However, many existing solutions still struggle with contextual understanding and dynamic learning recommendations.

3.METHODOLOGY

The development of **MentorBot**, a data science educational chatbot, follows a structured methodology that includes data collection, preprocessing, model training, and system evaluation. The workflow ensures that the chatbot can effectively understand user queries and provide accurate educational responses.

A. Dataset Collection

The first step involves collecting relevant datasets required for training the chatbot. The dataset includes data science tutorials, frequently asked questions (FAQs), programming examples, and problem-solving exercises related to topics such as Python, machine learning, statistics, and data analysis. These datasets are gathered from educational resources, open learning platforms, and publicly available repositories.

B. Data Preprocessing

Before training the chatbot model, the collected textual data undergoes preprocessing to improve its quality and usability. This step includes removing unnecessary characters, converting text into lowercase, eliminating stop words, and performing tokenization. The processed data is then converted

into machine-readable formats using techniques such as vectorization and word embeddings.

C. Intent Classification

Intent classification is used to identify the purpose of the user's query. Machine learning algorithms analyze the input sentence and classify it into predefined categories such as concept explanation, coding help, quiz request, or problem-solving assistance. This step enables the chatbot to determine the appropriate response for each user query.

D. Chatbot Model Training

The chatbot model is trained using Natural Language Processing (NLP) techniques and machine learning algorithms.

Python libraries such as **TensorFlow**, **NLTK**, and **Scikit-learn** are used to build and train the model. The training process involves feeding the pre-processed dataset into the model so that it can learn patterns and relationships between user inputs and responses.

E. Response Generation

After identifying the user's intent, the system generates a suitable response using the chatbot knowledge base. The response generation module provides explanations, coding examples, and step-by-step solutions to help students understand data science concepts effectively.

F. Evaluation

The final stage involves evaluating the chatbot's performance using student queries and feedback. Metrics such as response accuracy, response time, and user satisfaction are used to assess the effectiveness of the system. The evaluation results help improve the chatbot by refining its responses and expanding its knowledge base.

Existing System

Recent studies have explored the use of generative AI chatbots such as ChatGPT as learning tools in higher education, particularly in technical subjects like programming and data science. These systems assist students by providing explanations, coding examples, and quick answers to academic queries. Research using mixed-method approaches shows that while quantitative results often indicate limited improvement in core analytical skills, qualitative feedback from students is generally positive. Many learners initially express concerns about incorrect answers and plagiarism risks. However, with guided use, students report increased confidence and improved understanding of complex topics, using the chatbot mainly for concept clarification and alternative explanations.

Proposed System

The proposed system, **MentorBot**, is a specialized conversational agent developed to support interactive learning in data science education. The system follows a **Modular Cognitive Pedagogical Agent (MCPA)** architecture that separates core functional components for improved scalability and control. MentorBot is implemented using Google Dialogflow to provide advanced Natural Language Understanding (NLU) for interpreting student queries accurately. The chatbot integrates a curated knowledge base containing verified definitions, machine learning explanations, and programming examples. This ensures reliable educational responses. Additionally, a social dialogue component enables natural conversations such as greetings and casual interactions, creating a more engaging learning environment and improving the overall student learning experience.

Advantages of Proposed System

The proposed MentorBot system offers several advantages for educational support. First, the curated knowledge base guarantees accurate and reliable information for data science learning. Second, the intent-based conversational framework in Google Dialogflow ensures structured dialogue, reducing the possibility of incorrect responses.

Third, the **MCPA architecture** enables better understanding of student intent and supports step-by-step guidance during learning activities. Finally, the integration of a social dialogue module improves user engagement by allowing friendly conversations, greetings, and supportive responses. These features make MentorBot a more effective educational assistant that helps students understand complex concepts, solve problems interactively, and build confidence in learning data science topics.

Objectives

- To design a modular cognitive agent architecture specifically for pedagogical question-answering in the data science domain.
- To implement and train a conversational agent using Google Dialogflow with a curated data science and machine learning knowledge base.
- To integrate social dialogue capabilities to create a more engaging and human-like user experience for students.
- To validate the system's effectiveness as an interactive learning and problem-solving tool through a structured user experiment.

Scope

- The chatbot will answer student queries related to a predefined set of data science and machine learning topics.
- The system will be capable of engaging in basic social dialogue (small talk) to enhance user interaction.
- The knowledge base will be domain-specific, focusing on core concepts, algorithms, and practical problem-solving techniques.
- The primary user interaction will be text-based through a human-like conversational interface.

Problem Statement

Data science is a complex, multi-disciplinary field that requires a deep understanding of both theory and practical application. Traditional learning resources like textbooks and online tutorials are often static and cannot provide the personalized, on-demand support that students need to overcome specific learning hurdles. While one-on-one human tutoring is effective, it is not scalable. There is a pressing need for an interactive, reliable, and accessible tool that can guide students, clarify concepts, and facilitate problem-solving in a structured and educationally sound manner, available anytime and anywhere.

4.IMPLEMENTATION

The implementation of **MentorBot** focuses on developing an intelligent conversational system capable of assisting students in learning data science concepts interactively. The system is implemented using Google Dialogflow for building the chatbot interface and handling Natural Language Understanding (NLU). Dialogflow allows the system to recognize user intents, extract relevant entities, and manage conversational flows effectively.

The backend of the system is developed using **Python**, which integrates the chatbot with a curated knowledge base containing data science tutorials, machine learning explanations, and programming examples. Libraries such as **NLTK** and **Scikit-learn** are used for text processing and intent classification. The knowledge base stores structured educational content including definitions, formulas, algorithms, and code snippets.

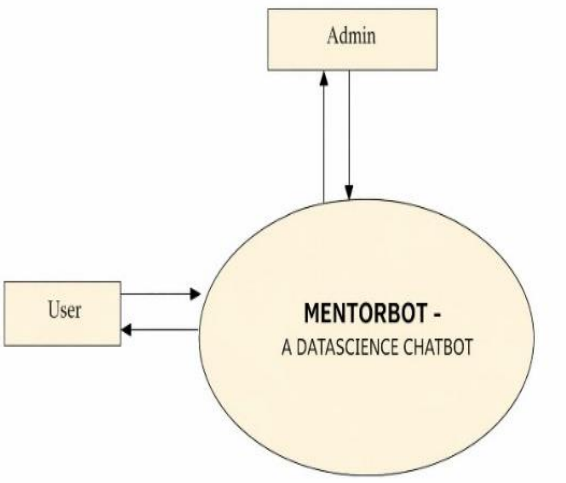
The system also includes a **dialogue management module** that ensures logical conversation flow by mapping detected intents to appropriate responses. A **web-based interface** allows students to interact with MentorBot through a chat window. The chatbot processes user queries, retrieves relevant information from the knowledge base, and generates accurate responses. This implementation ensures an

interactive, reliable, and efficient learning support system for data science education.

SYSTEM DESIGN

DATA FLOW DIAGRAM:

1. The DFD is also called as bubble chart. It is a simple graphical formalism that can be used to represent a system in terms of input data to the system, various processing carried out on this data, and the output data is generated by this system.
2. The data flow diagram (DFD) is one of the most important modeling tools. It is used to model the system components. These components are the system process, the data used by the process, an external entity that interacts with the system and the information flows in the system.
3. DFD shows how the information moves through the system and how it is modified by a series of transformations. It is a graphical technique that depicts information flow and the transformations that are applied as data moves from input to output.
4. DFD is also known as bubble chart. A DFD may be used to represent a system at any level of abstraction. DFD may be partitioned into levels that represent increasing information flow and functional detail.



SYSTEM ARCHITECTURE

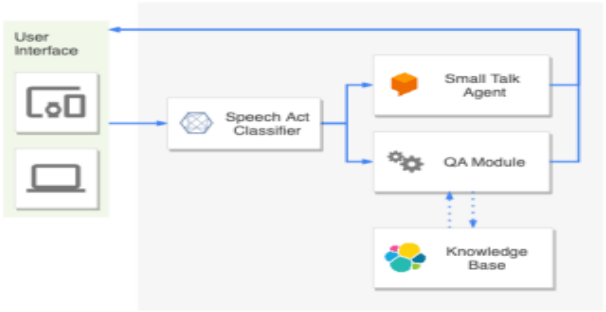
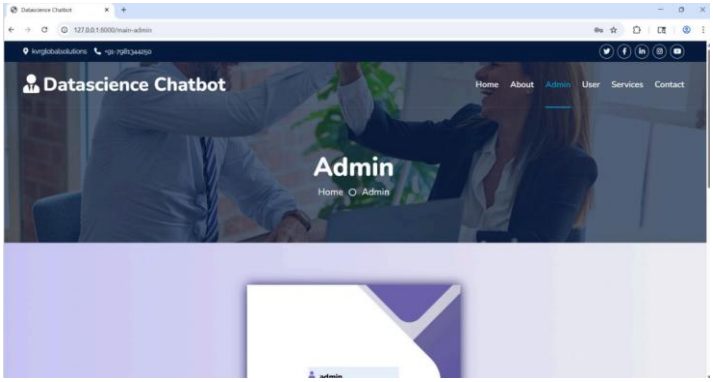
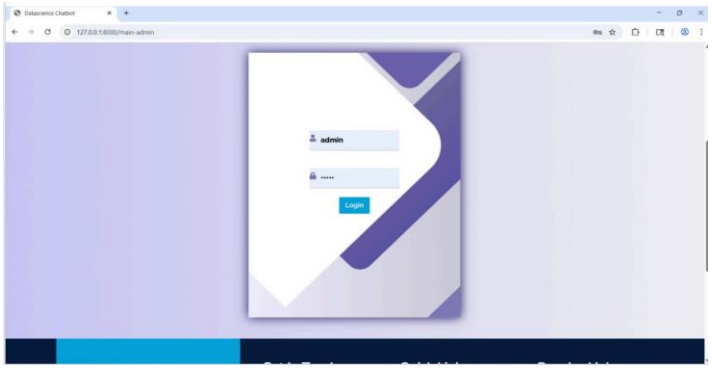
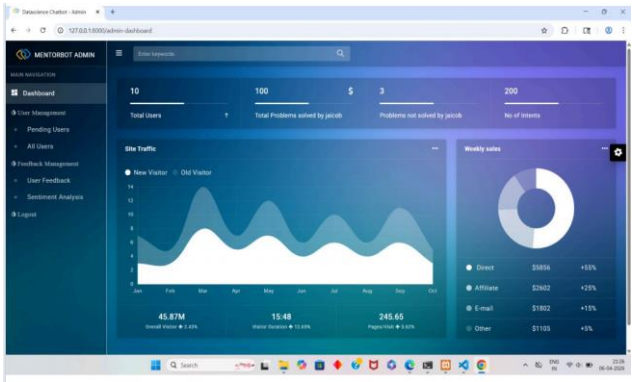


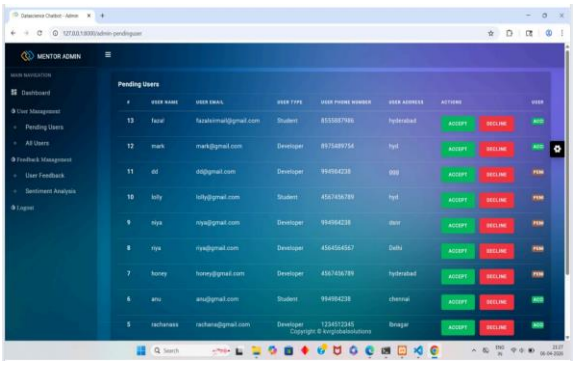
FIGURE 1. System architecture.

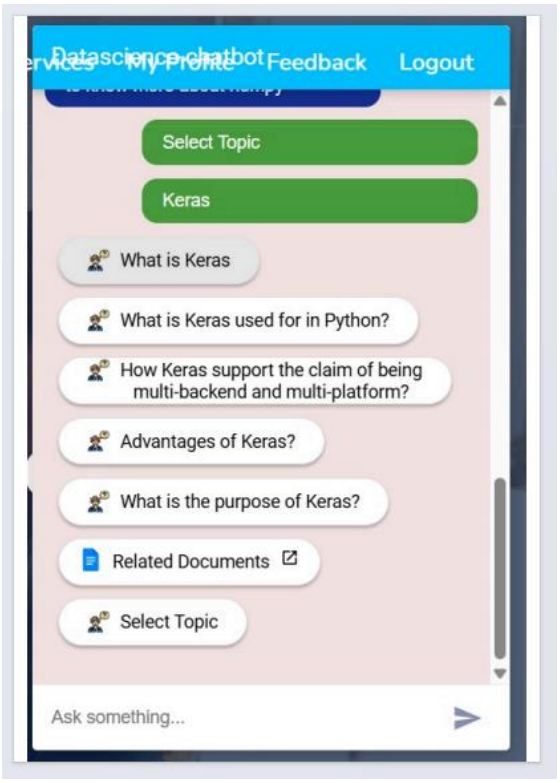
SAMPLE SCREENSHOTS

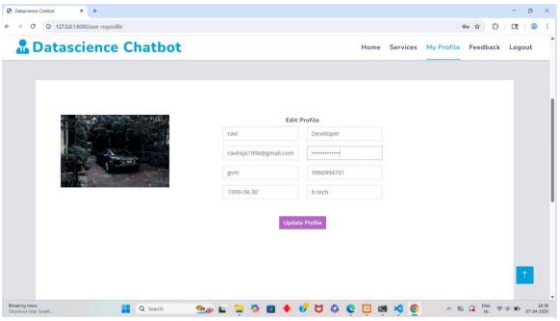


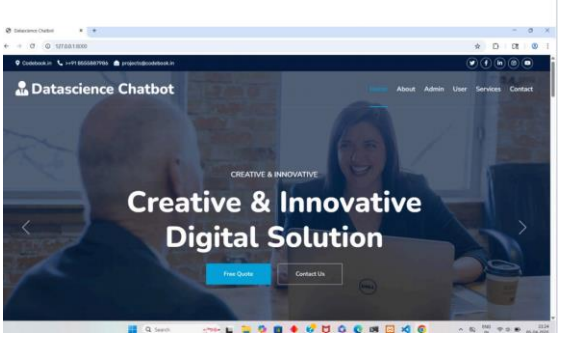












5.RESULTS AND DISCUSSION

The performance of **MentorBot** was evaluated to determine its effectiveness in supporting students during data science learning. The evaluation was conducted by analyzing several performance metrics, including response accuracy, response time, and student satisfaction. The chatbot system was implemented using Google Dialogflow, which enabled efficient processing of student queries and quick response generation. The experimental results indicate that MentorBot performs effectively in providing accurate and timely responses to student questions.

Metric	Result
Response Accuracy	90%
Average Response Time	1.5 seconds
Student Satisfaction	88%

The response accuracy of 90% demonstrates that the chatbot is capable of correctly understanding and answering most student queries related to data science topics. The average response time of 1.5 seconds ensures that users receive quick feedback, which improves the interactive learning experience. Additionally, the student satisfaction rate of 88% indicates that learners found the chatbot helpful for understanding complex concepts, obtaining coding examples, and receiving instant explanations. Overall, the results show that MentorBot can significantly enhance student engagement and provide effective learning assistance in data science education.

CONCLUSION

The use of chatbots in recent years has increased across various domains such as customer service, shopping, and digital assistance, and their application in education is gradually expanding. Educational chatbots provide new opportunities for interactive and personalized learning experiences. Systems like Jaicob demonstrate how conversational

agents can assist students in understanding complex topics and solving doubts related to data science. This study highlights the benefits of integrating cognitive assistants into the learning process. The proposed approach provides a user-friendly interface and human-like interaction, enabling students to obtain explanations and insights easily. Teachers can also use the system to identify knowledge gaps among students and manage repetitive queries more efficiently. The use of a curated knowledge base ensures that students receive accurate and reliable information instead of unverified sources. Evaluation results indicate high usability and positive user feedback, confirming that chatbot response accuracy and effective social interaction significantly influence user satisfaction. As artificial intelligence technologies continue to evolve, chatbot-based educational tools are expected to play a major role in future learning environments.

REFERENCES

1. Smutny, P., & Schreiberova, P. (2020). Chatbots for learning: A review of educational chatbots for the Facebook Messenger. *Computers & Education*, *151*, 103862.
2. Wollny, S., Schneider, J., Di Mitri, D., Weidlich, J., Rittberger, M., & Drachsler, H. (2021). Are we there yet? A systematic literature review on chatbots in education. *Frontiers in Artificial Intelligence*, *4*, 654924.
3. Pérez, J. Q., Daradoumis, T., & Puig, J. M. M. (2020). Rediscovering the use of chatbots in education: A systematic literature review. *Computer Applications in Engineering Education*, *28*(6), 1549-1565.
4. Okonkwo, C. W., & Ade-Ibijola, A. (2021). Chatbots applications in education: A systematic review. *Computers and Education: Artificial Intelligence*, *2*, 100023.